**Supplemental Experimental Procedures**


**Iterative stable alignment and clustering of 2D transmission electron microscope images**

Zhengfan Yang[a], Jia Fang[a], Johnathan Chittuluru[b], Francisco J. Asturias[b*], Pawel A. Penczek[a*]



[a]The University of Texas – Houston Medical School
Department of Biochemistry and Molecular Biology
6431 Fannin St, MSB 6.218, Houston TX 77030 USA

[b]Department of Cell Biology, The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla CA 92037 USA

**S1. The relation between Signal-to-Noise Ratio (SNR) and alignment stability for a homogenous data set**

In this test we used simulated data to demonstrate that, for a homogeneous set of images, the reference-free alignment initiated randomly is either successful or not, depending on the SNR of the data.  Put differently, for data with sufficiently high SNR, the alignment of a set of images is stable, i.e., the reference-free alignment will always succeed irrespective of initialization.  For low SNR values the alignment becomes unstable, i.e., the outcome of the reference-free alignment is essentially random.

We first generated a 2D projection image of the X-ray model of yeast RNA polymerase II (PDB accession code 1NT9) that had distinct features and was not pseudo-symmetric.  We then low-pass filtered this image using a Gaussian filter with a half-width of 0.3 absolute frequency units (0.5 corresponds to the Nyquist frequency).  Then we created its 100 noisy versions by corrupting it with Gaussian noise whose amplitude was adjusted to obtain the desired level of SNR and also shaped by an exponential Fourier filter $e^{-s/b}$, where $b$=2.0 and $s$ is a modulus of spatial frequency (Supplemental Figure S6).

For each test SNR value we applied the reference-free alignment algorithm five times using random initializations.  Based on the resulting alignment parameters we calculated the percentage of mirror stability (see Section IIId) and average group pixel error (see Section IIIc).

The relation between the alignment stability and SNR of the homogenous data set is almost binary (Supplemental Figure S7).  More specifically, the alignment is very stable if SNR is larger than 0.25 and it is very unstable if SNR is smaller than 0.1, and the transition zone is very narrow.

## S2. The relation between homogeneity and alignment stability

Here, in order to study the relation between group homogeneity and alignment stability, we selected two image groups resulting from ISAC processing of the EF-Tu ribosomal complex data set (see Section II(b)).  Based on the quality of the ribosomal data set and the high  purity of ISAC groups obtained from it, we assumed that each of the two groups selected was homogeneous (or sufficiently so for the purpose of this test).  We also created a third, heterogeneous set of images that contained half of the images from the first group and half from the second group.  We performed five independent reference-free alignments for each of the three data sets.  For each data set the averages are visually indistinguishable, which can lead to the incorrect conclusion that since they are the same in repeated alignment trials, the three sets have to be homogeneous (Supplemental Figure S8).  Similarly, the differences between Fourier Ring Correlation (FRC) curves computed for the aligned homogeneous and heterogeneous sets are minimal, so they cannot be used to determine which of the three groups is homogenous (Supplemental Figure S9).

Finally, we calculated the alignment stability for the homogeneous and heterogeneous groups using the five reference-free alignment results.  For mirror-stable particles, we sorted the pixel error of individual particles (Supplemental Figure S10).  For the two homogeneous groups all particles are mirror stable and the pixel errors are all small.  In contrast, for the heterogenous group, about 15% of the particles are mirror unstable and some have very large pixel errors, and are thus orientation-unstable.

We conclude that evaluation of alignment stability is a highly effective and sensitive tool for determination of homogeneity of set of EM particle images.  In addition, the alignment stability test makes it possible to detect a homogeneous subset of particle images within a larger, heterogeneous set.

## S3. Reproducibility *versus* homogeneity of a data set

The reproducibility of the ISAC results are expected to depend on the proper choice of the number of images per cluster in SAC.  To test this hypothesis we first selected 10 distinct 2D projection images of the discretized X-ray model of yeast RNA polymerase II.  We then created 200 noisy versions for each of these 10 images using the method described in Supplemental Section S1 above and by setting SNR = 1.0 and thus obtained a set of 2,000 test 2D images.

We first independently applied ISAC program twice to the test data set, each time setting the expected number of images per cluster to 200.  In both runs we were able to extract 10 clusters and the number of particles accounted for in respective runs was 1938 and 1946.  According to the reproducibility test, the number of matched particles was 1931 (97% of the total number) and all 10 clusters matched almost perfect.  Next, we again applied ISAC program twice, but with the expected number of images per cluster set to 100.  The first run yielded 20 clusters that accounted for 1,713 particles and the second run yielded 21 clusters that accounted for 1,655 particles.  However, according to the reproducibility test the number of matched particles was 1245, thus only 62% of the total number.

We conclude that the ability of ISAC to accomplish full reproducibility strongly depends on the correct setting of the expected number of images per cluster.  For experimental data set this number might be difficult to estimate correctly.   In addition, if the macromolecule adopts a preferential orientation and the data set is dominated by a particular view (or views), these very similar particle images will be distributed among a number of clusters.  If this is the case, the test results shown here demonstrate that a particle images are likely to change assignments between such clusters.  While this is by itself hardly consequential for subsequent steps data analysis, it will necessarily lower the accomplished reproducibility.

## S4. Comparison of ISAC with MRA results of the EF-Tu ribosomal complex data set analysis

In order to better understand the properties of the ISAC approach, we compared the ISAC results with results from MRA analysis of the same test data set of 50,000 projection images of EF-Tu ribosomal complex.   For fairness sake, we set the number of clusters in MRA to 632 (50000/79=632), so the expected average number of particles per cluster should be the same as in ISAC.   However, MRA yielded only 349 clusters that had more than one member (Supplemental Figure S11), the other clusters collapsed.  We found that only 202 of these 349 clusters were alignment stable, i.e., had average group pixel error lower than $\sqrt{3}$ pixels.  This indicates that the ISAC approach, which yielded a larger number of clusters (471) that are alignment stable, performs clearly better than MRA.

To compare the homogeneity of the 471 ISAC clusters to that of the 202 MRA clusters that are alignment stable, we performed projection matching of ISAC and MRA averages using the published 6.4 Å map of EF-Tu ribosomal complex as a template and found projection angles for all cluster averages.  For each cluster, we calculated the deviation of projection angles assigned to individual EM projection images in the structure determination process that results in 6.5A map of EF-Tu ribosomal complex from the projection angle of this cluster average.   Finally, we

computed average deviations of projection angles for all images within each cluster.  e found that for ISAC clusters the median value of average deviations of projection angles was 5.2 degrees.  In contrast, for MRA clusters that are alignment stable this value was 20.0 degrees , thus nearly four times larger,.  We further conducted a similar test in which we performed projection matching of all 50,000 individual particles as well as ISAC and MRA cluster averages using the X-ray model of EF-Tu ribosomal complex as a template.  We found that for ISAC clusters the median value of average deviations of projection angles was 5.9 degrees and for MRA clusters that are alignment stable the median value was 20.4 degrees.  Taken together, results of these two tests provide very strong evidences that ISAC clusters are more homogeneous than those obtained using MRA.

Finally, we compared the reproducibility of the ISAC and MRA algorithms by applying them a second time to the test 70S ribosome data set using differently randomized initial settings.  We compared the results from two independent runs of each algorithm and computed reproducibility by setting the minimum size of clusters accepted for matching to 20 (see Section III (e)). To evaluate how well two clusters match, we used a notion of matching percentage, defined as the number of images shared by two clusters divided by the mean size of two clusters.  When two clusters are the same (i.e., perfectly matched), the matching percentage is 100%; when two clusters have no intersection, it is 0%.  For the two independent runs of ISAC, the reproducibility test yielded 175 matches for clusters larger or equal than 20.   The average matching percentage for all 175 reproduced matches of ISAC clusters was 45%.   The comparison between the two independent MRA runs yielded 125 matches (for clusters containing more than 20 members), and the average matching percentage was 36%.  We conclude that ISAC clusters are more reproducible than MRA clusters by a large margin, although they fall short of full reproducibility.

The lack of full reproducibility for ISAC results can be attributed to the properties of the test ribosomal data set, namely the fact that due to strongly preferential orientation of the complex ISAC delivered many very similar averages.  As the respective clusters necessarily comprise similar images, their reassignment to similar clusters have limited effect on the stability tests used within ISAC to determine cluster homogeneity.
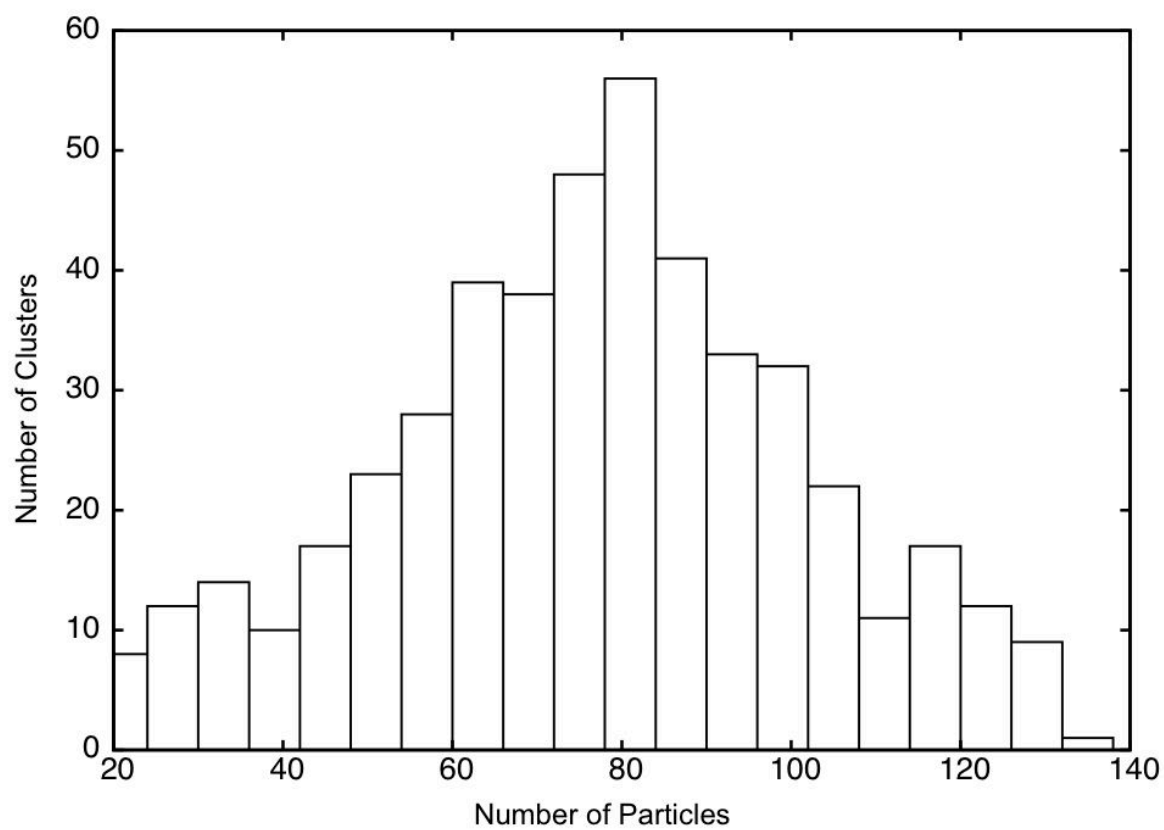
# Supplemental Figures



Figure S1.   Histogram of the number of particle images per cluster as obtained from ISAC analysis of the EF-Tu data set.
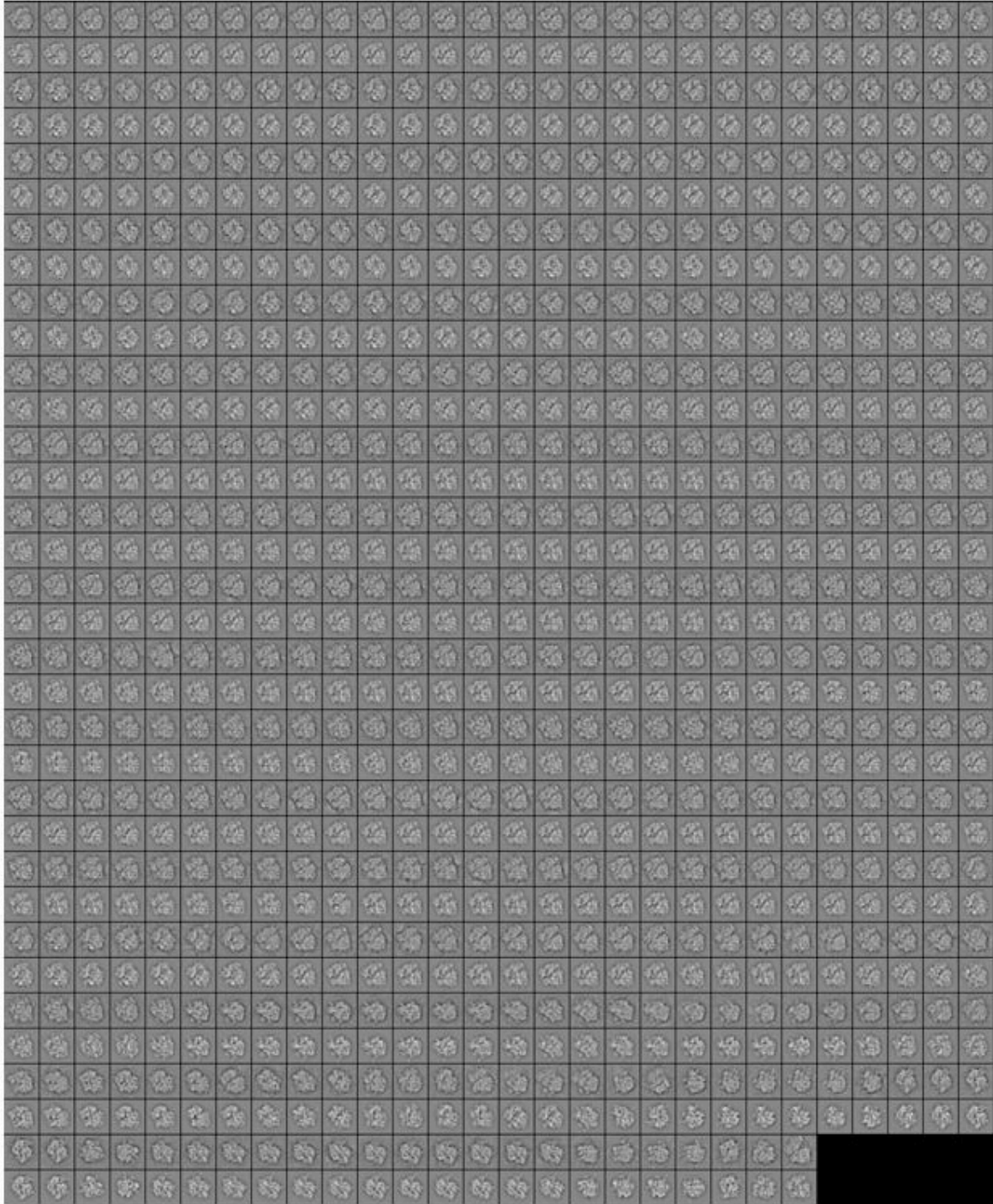
Figure S2. Set of 471 ISAC cluster averages obtained from analysis of 50,000 particle images of the EF-Tu ribosomal complex and matching reprojections of the X-ray model. The ISAC averages were mutually aligned and ordered according to pair-wise similarity. Power spectra of reprojections were adjusted to match the average of rotational power spectrum of EM averages. The ISAC cluster averages are shown in odd rows and their matching reprojections are shown in even rows, counting from the top of the panel.
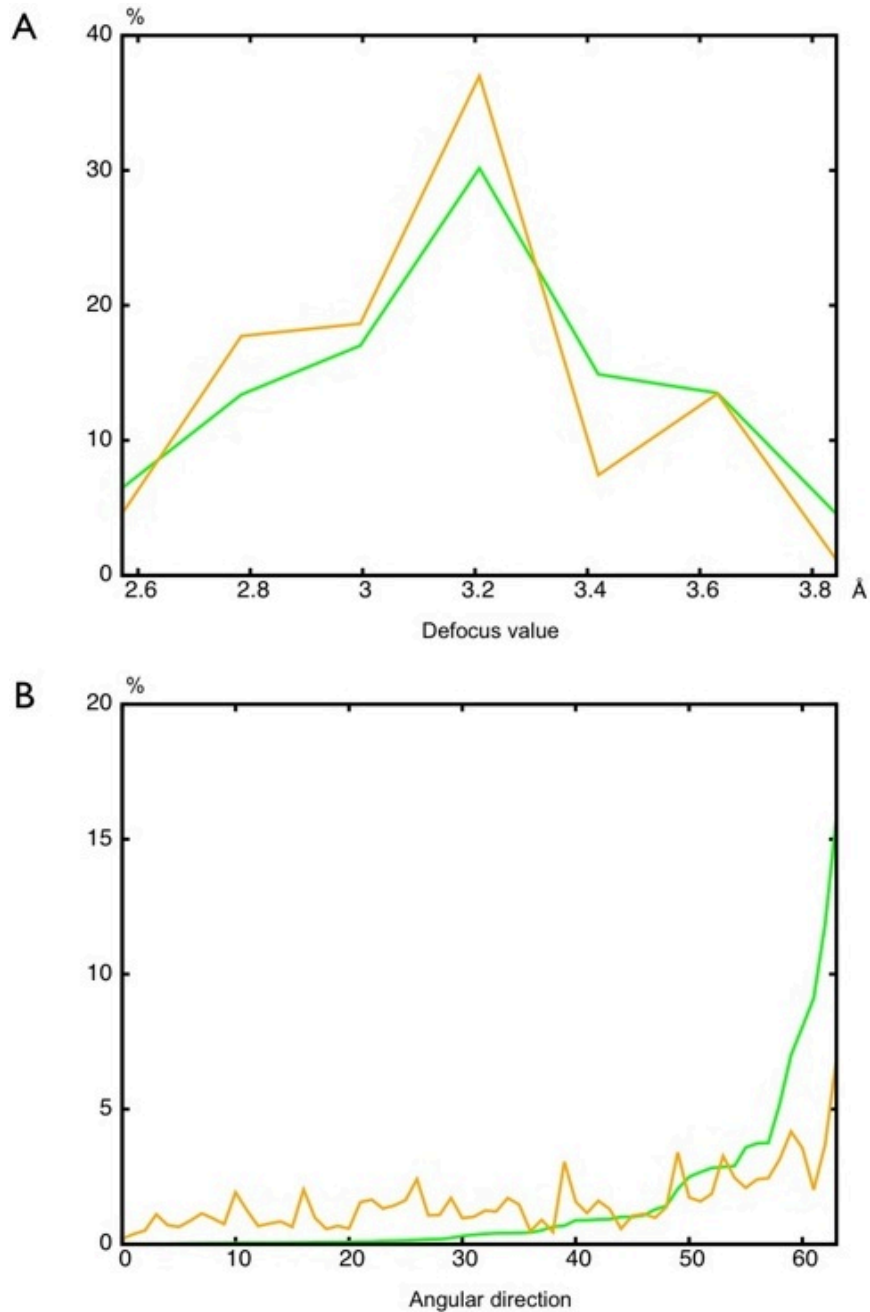
Figure S3. The comparison of defocus values and angular distributions of images accounted for (green) and unaccounted for (orange) in ISAC analysis of 50,000 EF-Tu data set. (A) Histogram of a number of projection images as a function of assigned defocus values. (B) Histogram of an angular distribution of EF-Tu projection images, as assigned by projection matching to X-ray model. In each plot, the y-axis represents percentage of images within respective set, i.e., accounted and unaccounted for.
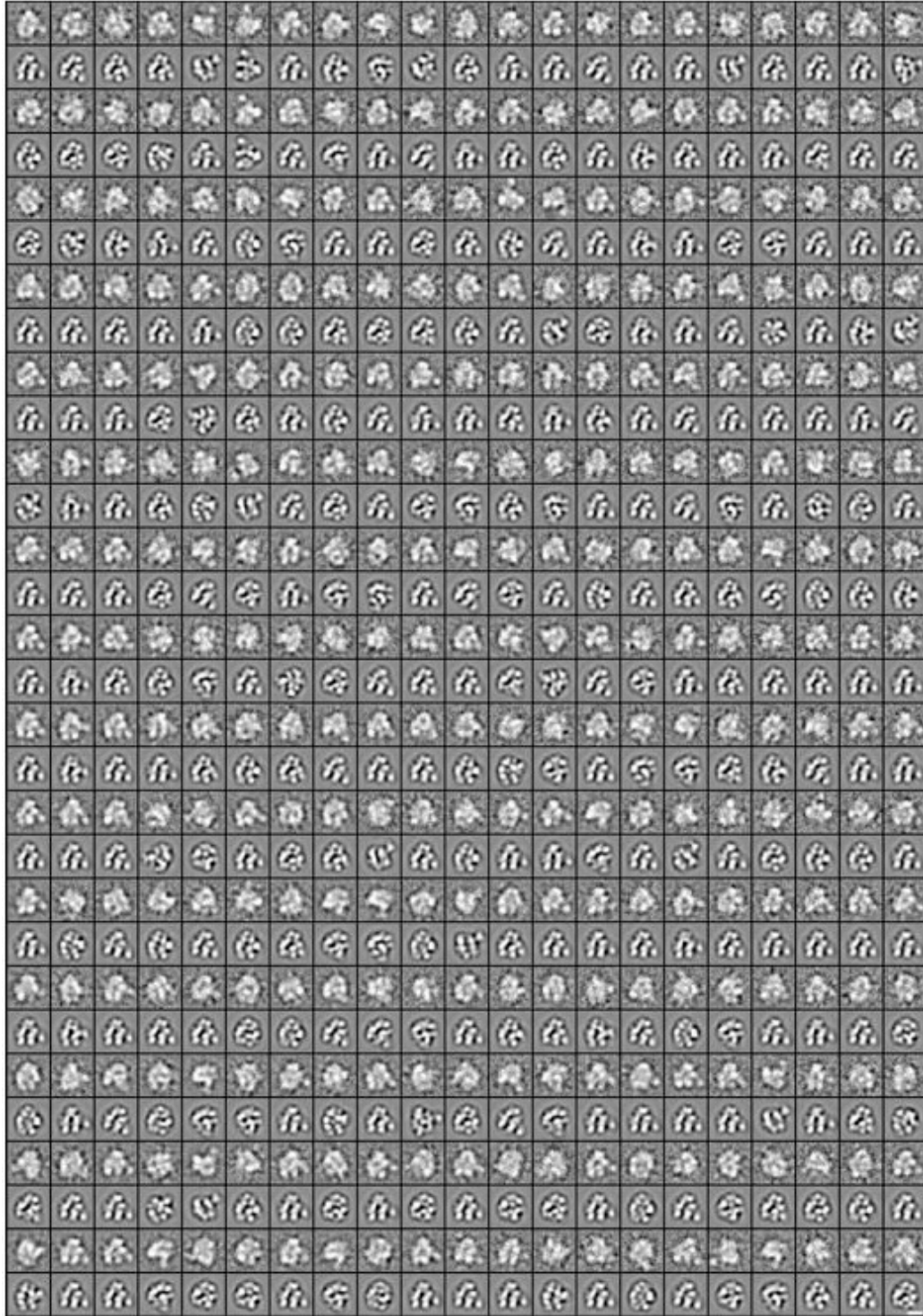
Figure S4. hRNAPII ISAC cluster averages and comparison with reprojections of the X-ray structure of the homologous yeast RNAPII. Power spectra of reprojections were adjusted to match the average of rotational power spectrum of EM averages. The ISAC cluster averages are shown in odd rows and their matching reprojections are shown in even rows, counting from the top of the panel.

Figure S5. The ISAC GUI of the SPARX system. (A) The main panel for setting major ISAC parameters, retrieving and saving input parameters, generating command line from input parameters and running ISAC program and (B) panel for advanced parameters.
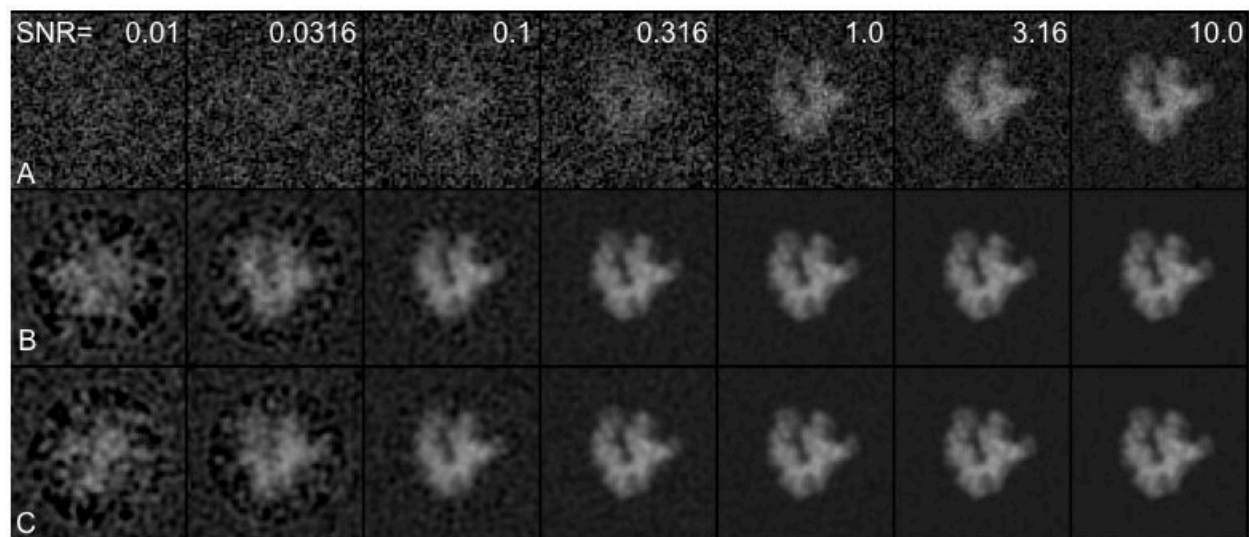
Figure S6. The relation between the alignment stability and SNR of the homogenous data set of 100 noise-corrupted simulated images. (A) Examples of individual simulated particle images corrupted by Gaussian noise and their SNR. (B-C) Averages of the aligned data set resulting from two independent trials.
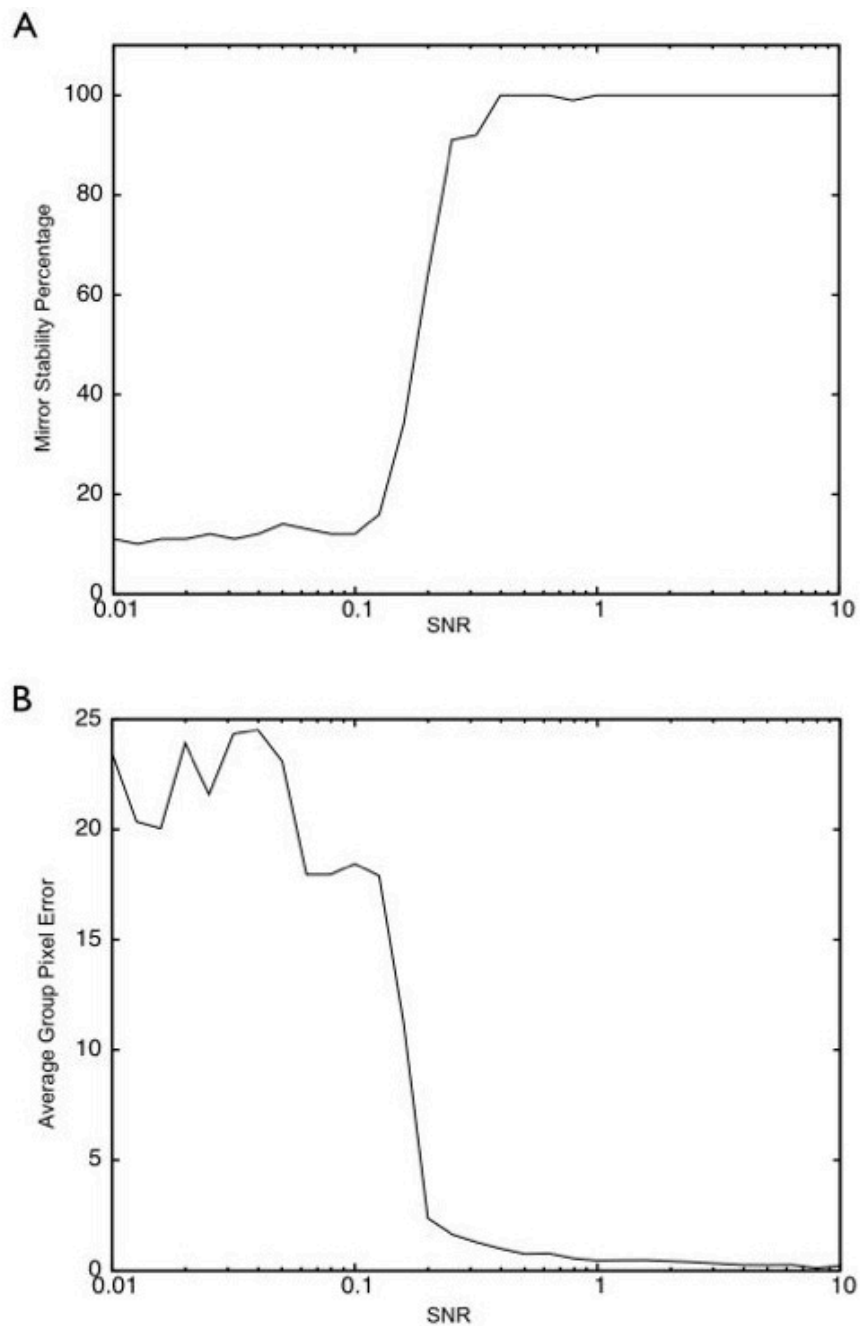
Figure S7. Two indicators of alignment stability of a homogeneous data set as a function of SNR of the data set. The test data sets comprised of 100 images created using a projection image of a discretized atomic model of yeast RNA polymerase II corrupted by Gaussian pink noise. The results were computed based on 5 independent reference-free alignments for each tested level of SNR of the data. (A) Percentage of mirror stable images and (B) average pixel error (Eq.3).
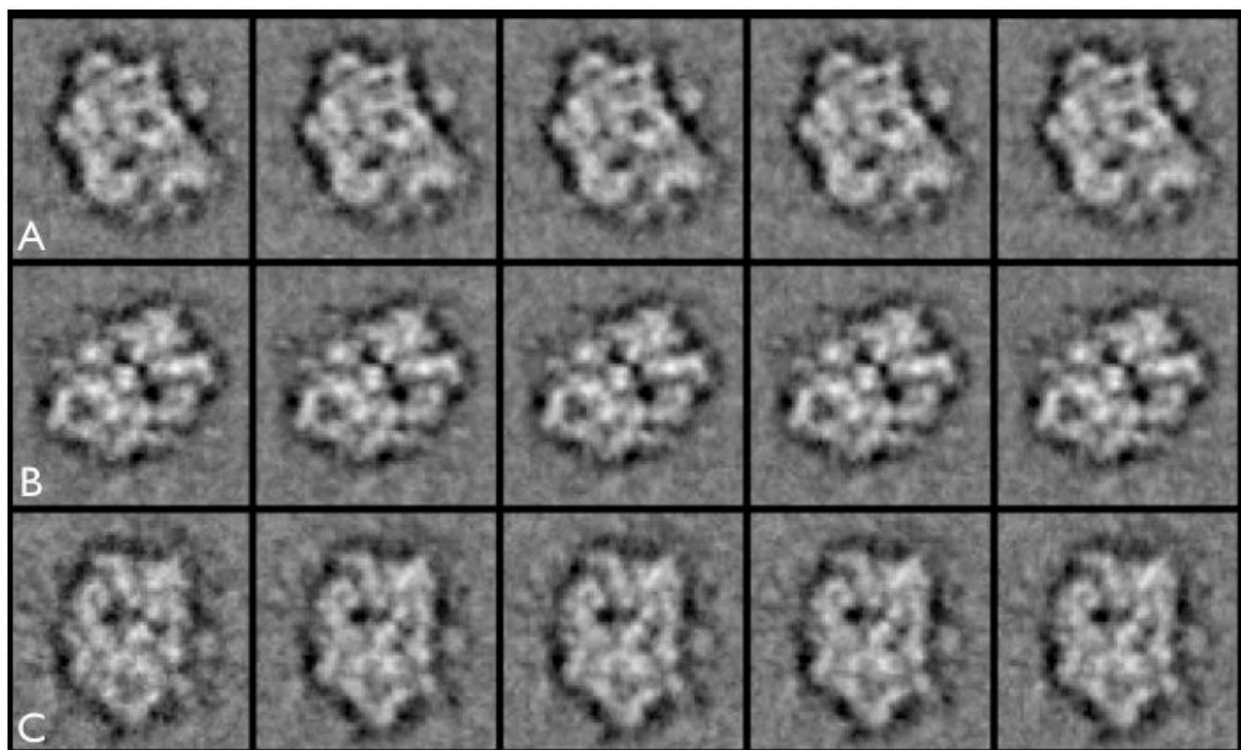
Figure S8. Reference-free alignment averages of image groups can be similar irrespective of the homogeneity of the group. (A-B) Averages resulting from five independent reference-free alignments of the two nearly homogeneous image groups drawn from ISAC processing of the EF-Tu ribosomal complex data set. (C) Averages resulting from five reference-free alignments of the heterogeneous class created by randomly selecting half of the particles from the first group and another half from the second group.
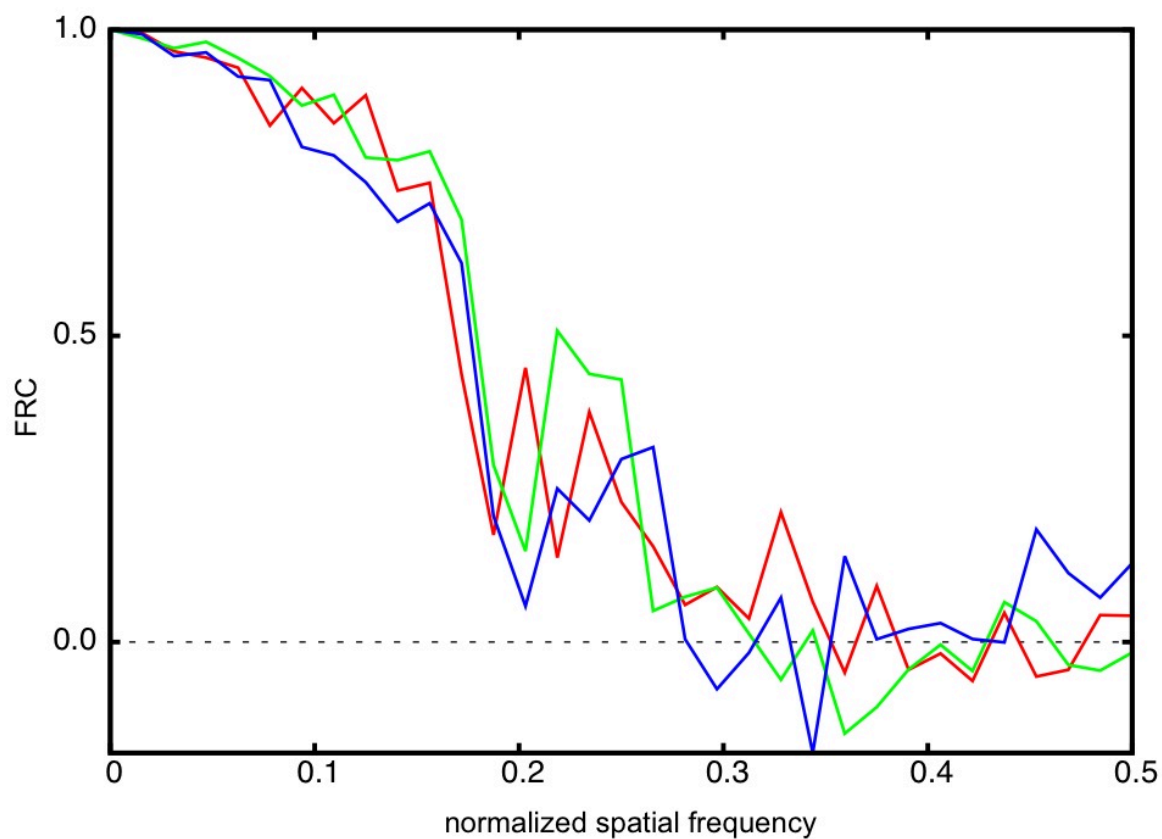
Figure S9. Fourier Ring Correlation (FRC) curves computed for selected alignments of the homogeneous (green and red lines) and heterogeneous (blue line) sets of selected groups of EF-Tu images whose averages are shown in Figure S8.
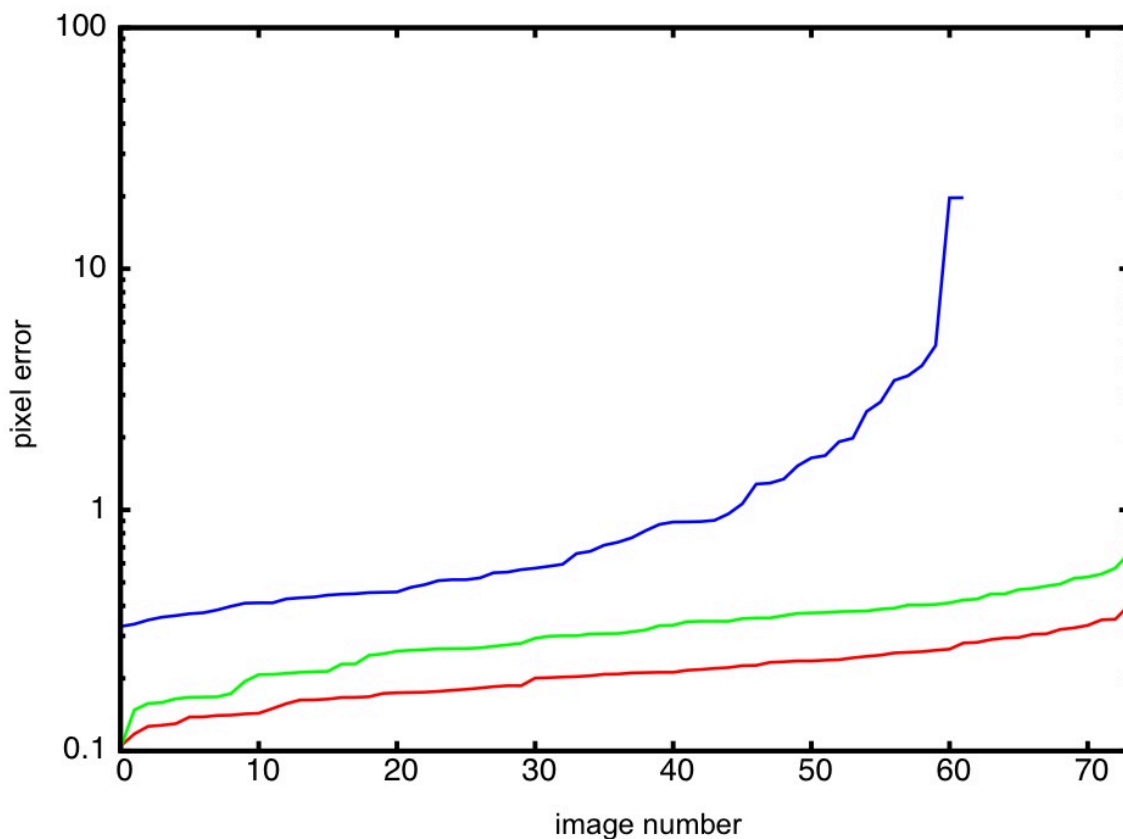
Figure S10. Pixel errors of individual particle images computed using the alignment stability test of homogenous (green and red lines) and heterogeneous (blue line) groups. The alignment stability test is done using results of five independent reference-free alignments of selected groups of EF-Tu images whose averages are shown in Figure S8. The missing 'tail' of the blue line corresponds to the mirror-unstable images, for which we do not evaluate pixel errors. The images were sorted by increasing values of their pixel errors.
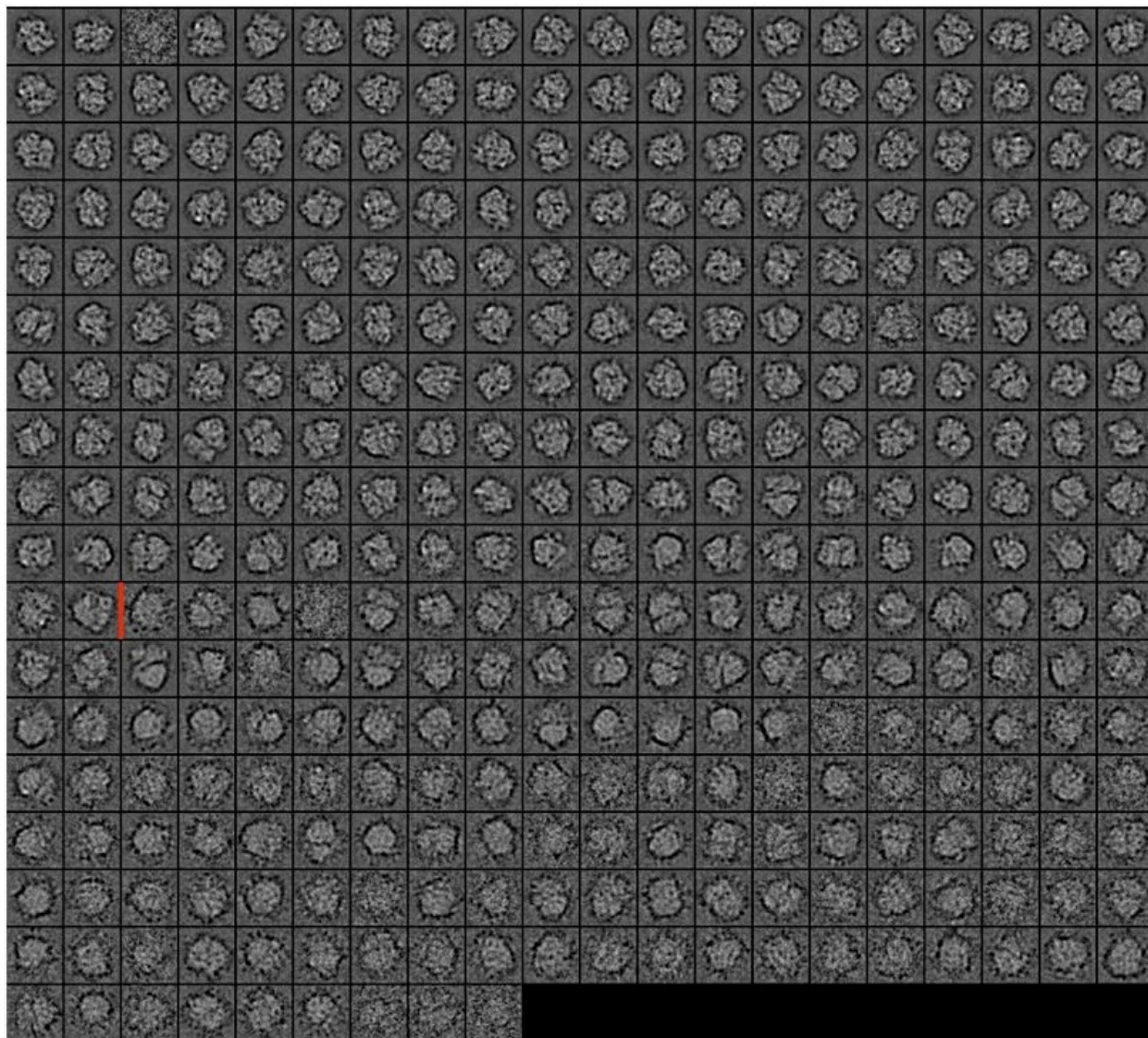
Figure S11.　349 MRA cluster averages (containing more than 1 image) of EF-Tu data set arranged in the ascending order of average group pixel error.　The red line indicates the boundary between alignment stable cluster averages and those that are alignment unstable. Pixel error threshold was set to $\sqrt{3}$ .